PERGAMON

Neural Networks 12 (1999) 281–297

Neural Networks

Contributed article

Improved bidirectional retrieval of sparse patterns stored by Hebbian learning

Friedrich T. Sommer*, Günther Palm

Department of Neural Information Processing, University of Ulm, 89069 Ulm, Germany

Received 26 June 1997; received in revised form 11 September 1998; accepted 11 September 1998

Abstract

The Willshaw model is asymptotically the most efficient neural associative memory (NAM), but its finite version is hampered by high retrieval errors. Iterative retrieval has been proposed in a large number of different models to improve performance in auto-association tasks. In this paper, bidirectional retrieval for the hetero-associative memory task is considered: we define information efficiency as a general performance measure for bidirectional associative memory (BAM) and determine its asymptotic bound for the bidirectional Willshaw model. For the finite Willshaw model, an efficient new bidirectional retrieval strategy is proposed, the appropriate combinatorial model analysis is derived, and implications of the proposed sparse BAM for applications and brain theory are discussed. The distribution of the dendritic sum in the finite Willshaw model given by Buckingham and Willshaw [Buckingham, J., & Willshaw, D. (1992). Performance characteristics of associative nets. *Network*, *3*, 407–414] allows no fast numerical evaluation. We derive a combinatorial formula with a highly reduced evaluation time that is used in the improved error analysis of the basic model and for estimation of the retrieval error in the naive model extension, where bidirectional retrieval is employed in the hetero-associative Willshaw model. The analysis rules out the naive BAM extension as a promising improvement. A new bidirectional retrieval algorithm — called crosswise bidirectional (CB) retrieval — is presented. The cross talk error is significantly reduced without employing more complex learning procedures or dummy augmentation in the pattern coding, as proposed in other refined BAM models [Wang, Y. F., Cruz, J. B., & Mulligan, J. H. (1990). Two coding strategies for bidirectional associative memory. *IEEE Trans. Neural Networks*, 1(1), 81–92; Leung, C.-S., Chan, L.-W., & Lai, E. (1995). Stability, capacity and statistical dynamics of second-order bidirectional associative memory. *IEEE Trans. Syst. Man Cybern.*, 25(10), 1414–1424]. The improved performance of CB retrieval is shown by a combinatorial analysis of the first step and by simulation experiments: it allows very efficient hetero-associative mapping, as well as auto-associative completion for sparse patterns — the experimentally achieved information efficiency is close to the asymptotic bound. The different retrieval methods in the hetero-associative Willshaw matrix are discussed as Boolean linear optimization problems. The improved BAM model opens interesting new perspectives, for instance, in information retrieval it allows efficient data access providing segmentation of ambiguous user input, relevance feedback and relevance ranking. Finally, we discuss BAM models as functional models for reciprocal cortico–cortical pathways, and the implication of this for a more flexible version of Hebbian cell-assemblies.

Keywords: Bidirectional associative memory; Hebbian learning; Iterative retrieval; Combinatorial analysis; Cell-assemblies; Neural information retrieval

Nomenclature

a	number of active components in an x-memory
A	output capacity describing hetero-association
b	number of active components in a y-memory
C	completion capacity describing auto-association
C_{ij}	memory matrix formed by clipped Hebbian learning
e	number of hits obtained with pattern part retrieval
f	number of components extracted by pattern part retrieval
$\mathscr{F}$	set of fixed point pattern pairs of the memory matrix C_{ij}
g	number of 'false alarm' bits in an initial x-pattern
$H(x)$	Heavyside function
i	index of x-pattern component (subscript)
$i(p)$	Shannon information of a binary event with probability p

* Corresponding author. Tel.: + 49-731-502-4154; Fax: + 49-731-502-4156; e-mail: Friedrich.Sommer@informatik.uni-ulm.de

PII: S0893-6080(98)00125-7

$i(p,\alpha,\beta)$	conditional information in a noisy channel with error probabilities α and β for a binary event with probability p
j	index of y-pattern component (subscript)
k	index of x-pattern component (subscript)
l	index of y-pattern component (subscript)
m	length of the y-patterns
M	number of learning pattern pairs
$\mathcal{M}$	association mapping $x^\nu \rightarrow y^\nu$ performed by the memory
n	length of the x-patterns
p	ratio between active and passive components in a x-memory
$P_B(p,c;d)$	binomial distribution
$P_D(x_1,\cdots,x_n,c;d)$	distribution of integer variable d with $0 \le d \le c, x_1,\cdots,x_n$ additional parameters, D descriptor of the distribution
$p_r(z)$	hit probability for pattern part retrieval
p_1	probability that a memory matrix element is one
q	ratio between active and passive components in a y-memory
$Q_D(x_1,\cdots,x_n,c;d)$	$=\sum_{i=d}^{c} P_D(x_1,\cdots,x_n,c;d)$ cumulative distribution
r	index of the iteration time step
s	index of the iteration time step
S	search capacity describing bidirectional hetero-association
$\mathcal{S}$	set of memory pattern pairs
$t(p,\alpha,\beta)$	transinformation of a noisy channel with error probabilities α and β for a binary event with probability p
u	unit usage index
$w_{il}^{x(r)}$	conditional link between y units i,l for given pattern $x(r)$
$w_{jk}^{y(r)}$	conditional link between x units j,k for given pattern $y(r)$
x	binary pattern of length n
X	binary pattern of length n
$\tilde{x}^\mu$	noisy version of the memory x^μ used as initial pattern
$\|x\|$	number of active components in pattern x
y	binary pattern of length m
Y	binary pattern of length m
$\hat{y}$	output pattern of the standard Willshaw model
z	number of 'miss' bits in an initial x-pattern
$\alpha(r)$	retrieval 'false alarm' error probability in $y(r)$-patterns
$\beta(r)$	retrieval 'miss' error probability in $y(r)$-patterns
$\gamma(r)$	retrieval 'false alarm' error probability in $x(r)$-patterns
γ^{I}	'False alarm' error probability in initial patterns $x(0)$
$\delta_{m,n}$	Kronecker symbol
$\zeta(r)$	retrieval 'miss' error probability in $x(r)$-patterns
ζ^{I}	'Miss' error probability in initial patterns $x(0)$
$\Theta(r)$	neural threshold of the y-units in time step r
μ	index of memory patterns (superscript)
ν	index of memory patterns (superscript)
$\Xi(r)$	neural threshold of the x-units in time step r

1. Introduction

In the late fifties, Steinbuch (1961) proposed one of the first hardware realizations of a neural associative memory (NAM) model, his so called 'Lernmatrix', where binary synapses are formed by Hebbian local learning from binary memory patterns and hetero-associative one-step retrieval is performed. This model is now referred to as the Willshaw model, since Willshaw et al. (1969) first determined its asymptotic[1] information efficiency as $\ln[2] = 0.69$ bit/synapse. Such high efficiencies are only achieved in models with sparse memory patterns where the ratio between active and passive elements is far below 0.5 (Palm and Sommer, 1995), and asymptotically, the Willshaw model is the most efficient NAM, see the model comparison in Sommer and Dayan (1998). Unfortunately, at maximum memory load, the finite Willshaw network retrieves with high error rates (see Section 5). Error reduction in the Willshaw model can only be achieved by reducing the memory load that leads to information efficiencies far below the asymptotic value. For the Hopfield task, namely auto-associative pattern completion (Hopfield, 1982), iterative retrieval has been introduced in the Willshaw model (Gardner-Medwin, 1976; Gibson and Robinson, 1992; Hirase and Recce, 1996; Schwenker et al., 1996). For auto-association of finite sparse patterns, it is now well understood that the original retrieval process is the limiting bottleneck (Palm and Sommer, 1992), and what kind of retrieval modifications are most promising in the light of probabilistic reasoning (Sommer and Dayan, 1998).

This paper considers bidirectional retrieval in the hetero-associative Willshaw model. Hetero-associative iterative retrieval schemes have been proposed in bidirectional associative memory (BAM) models for dense patterns (Kosko, 1987), where two sets of threshold units are bidirectionally connected by a synaptic weight matrix. An improved theoretical analysis for the Willshaw model is derived and applied to the straightforward BAM extension of the Willshaw model (Haines and Hecht-Nielsen, 1988), referred to as standard bidirectional (SB) retrieval in the following, where standard retrieval is performed alternatingly for the two disjunct sets of units, the x and the y layers. By this analysis, the SB model can be ruled out as a promising model variant. Standard retrieval discriminates active neurons by thresholding the overlap between the vector of

[1] i.e. for infinitely large systems.

synaptic values and the activity pattern in the other layer. We propose a completely new retrieval method, crosswise bidirectional (CB) retrieval, based on dynamic virtual connections between units within a layer called conditional links. The conditional link between two neurons depends on the activity pattern in the other layer: it is defined by the overlap between the parts of their synaptic vectors that receive input from active units in the other layer. By the structure of the Hebbian synaptic matrix in the Willshaw model, the conditional link has a probabilistic interpretation: a high value indicates a high probability that the corresponding unit pair belongs to the memory pattern that should be associated with the activity pattern in the other layer. CB retrieval employs bidirectional activity propagation to determine the clique of units that are connected by the highest conditional links. CB retrieval uses dendritic sums weighted by conditional links that can be computed in parallel — by bidirectional propagation through the synaptic matrix. Thus, each CB update step involves evaluation of columns and rows as well, a fact that led to the name of this retrieval strategy. The advantage of CB retrieval is shown by the analysis of a single step and by simulation experiments.

The paper is organized into six sections. Section 2 revisits briefly the Willshaw NAM, its biological motivation and some basic definitions; it also explains the relations and differences between the hetero- and auto-association task and what memory models can combine both functionalities. In Section 3, we define bidirectional extensions of the model, namely SB retrieval — the classical BAM scheme, and CB retrieval based on conditional links. We present two versions of CB retrieval that differ in the iteration scheme and in the means to limit activity in the network. The theoretical background of the proposed retrieval strategies is developed in Section 4. In Section 4.1, the information efficiency is defined as a general performance measure for BAM and in Section 4.2, the asymptotic efficiency bound for the Willshaw model is determined. Section 4.3 contains a refined analysis of the finite size Willshaw model: an improved combinatorial calculation of retrieval error probabilities is derived (Proposition 4.2) that allows much faster numerical evaluation than the previously given formula (Buckingham and Willshaw, 1993). We analyze pattern part retrieval (Proposition 4.3), a method using the standard model to extract a part of the 1-elements in the memory pattern with higher accuracy. The analysis of the standard model is used in Section 4.4 to estimate the retrieval error of SB retrieval (Proposition 4.4). Section 4.5 analyzes the first step of CB retrieval (Proposition 4.5). Section 4.6 identifies the problem of optimal retrieval (POOR) as a Boolean linear optimization problem. In this framework, we point out the differences between SB and CB retrieval in Section 4.7. Section 5 presents some typical simulation results with the CB retrieval method and compares it with the Willshaw model. The conclusion section resumes the implications of the presented analysis (Section 6.1) and the properties of the proposed new BAM model (Section 6.2). The perspectives of applying the CB model in information retrieval are discussed in Section 6.3. We close with some speculations about the biological realization of BAM in reciprocal cortico–cortical pathways and the impact of such a cortex model on concept formation and processing (Section 6.4).

2. Neural associative memory (NAM) models

2.1. Motivations for associative memory models

In one of the first attempts of computational brain modeling, McCulloch and Pitts (1943) proposed a neural network of binary ‘all or none’ units and showed the computational universality of such systems. The psychologist Hebb (1949) speculated that psychological concepts could be represented by simultaneous activity of many nerve cells distributed throughout the brain, which he called a cell-assembly. He postulated that cell-assemblies are formed by an amplification process taking place in all synapses between active nerve cells during learning. This process of synaptic strengthening depending on coincident neural activity in the direct vicinity of the synapse was later called Hebbian learning. The simplest class of neural associative memories (NAM) are mathematical models of Hebb’s idea: in such models, activity patterns (of nerve cells) are stored in a matrix of synaptic connections during a one-step storage process employing a local synaptic learning rule (see Section 2.2) (Willshaw et al., 1969; Palm, 1980; Hopfield, 1982). Following Hebb, a learned assembly can be recalled if enough of its assembly neurons become activated. The corresponding recall process in a NAM is the retrieval process.

NAM models have been proposed for efficient searching in large data-bases allowing fast, fault-tolerant access and being particularly well-suited for parallel implementations (Steinbuch, 1961; Kohonen, 1977; Bentz et al., 1989). An appropriate performance measure for NAMs is the information efficiency, the amount of information that can be stored (and recalled) per bit of synaptic memory (see Section 4.2). The search for models with high information efficiency is one essential issue towards an application of NAM models.[2] Other important issues are sparse coding prescriptions for particular data, (see Section 6.3), and efficient Hardware implementations of NAM (Potter, 1992; Palm et al., 1997).

2.2. The Willshaw model

Consider the computational task where a mapping $\mathcal{M}: x^\nu \to y^\nu$ (for $\nu = 1,\cdots,M$) has to be performed between binary patterns. We call the pattern pairs involved the memory patterns, or simply the memories: $\mathcal{S} := \{(x^\nu, y^\nu): x^\nu \in \{0,1\}^n, y^\nu \in \{0,1\}^m, \nu = 1,\cdots,M\}$. The number of

[2] Nevertheless, information efficiency has also been proposed as a guideline for biological modeling by Palm (1990).

active ('one') components in a pattern is called pattern activity: $|y|:=\sum_{i=1}^{n} y_i$. The pattern symbol y is also identified with the set of its active components: $y=\{i{:}y_i=1\}$. Consequently, a pattern y' is called part of y if $y' \subset y$ holds. In the following, we use memories with constant activities, i.e. $|x^\nu|=a$, $|y^\nu|=b\ \forall\nu$. Noise in a memory is expressed by the two possible error types that can change a component: a 'miss' error converts a 1-entry to '0' and a 'false alarm' or 'add' error does the opposite.

During learning, the set of memories is transformed into the synaptic weight matrix:

$$C_{ij}=\min\left(1,\sum_\nu x_i^\nu y_j^\nu\right)=1-\prod_\nu\left(1-x_i^\nu y_j^\nu\right)=\sup_\nu x_i^\nu y_j^\nu \tag{1}$$

Note that this learning process provides distributed storage, i.e. learning of one memory trace affects many synapses and one synapse is in general affected from traces of several memories.

During retrieval, an initial pattern is mapped to its corresponding memory. The retrieval process is called an association, if the mapping is fault tolerant, in the sense that its domain also contains noisy memories. For a given initial pattern $\tilde{x}^\mu$, one-step retrieval yields the output pattern $\hat{y}$, by determining the dendritic sum in each neuron as the overlap between the input and the synaptic vector:

$$[C\tilde{x}^\mu]_j=\sum_i C_{ij}\tilde{x}_i^\mu \tag{2}$$

and by calculating the activity value by threshold comparison:

$$\hat{y}_j^\mu=H([C\tilde{x}^\mu]_j-\Theta)\ \forall j \tag{3}$$

with a global threshold value Θ, and $H(x)$ denoting the Heavyside function, to be evaluated for each vector component.

In fault-tolerant NAM models with distributed storage the retrieval output can also contain noise. Successful retrieval, i.e. a low Hamming distance to the desired memory y^μ, depends on the following conditions that have to be fulfilled:

- The stored patterns are sparse, i.e. the activities of the memories are similar and of the order of the logarithm of the pattern length.
- The 1-elements in the memories have even distribution over the pattern components.
- The number of memories is below a certain limit M^*, i.e. cross talk due to the superposition of different memory traces is kept small enough.
- The error level in the initial pattern is sufficiently low.
- The threshold is adjusted properly, that is, $\Theta\simeq\sum_i \tilde{x}_i^\mu x_i^\mu$.

The given threshold setting we call 'no misses' threshold setting because dendritic sums at 1-elements will assume or exceed this value and therefore the retrieved pattern contains no 'miss' errors at all. There is a trade-off between the limits on the model performance implied by the listed conditions: for instance, if the cross talk is increased by a higher memory load M, fault tolerance, that is, the limit on the admitted initial error level decreases. The most unpleasant property of the finite Willshaw model is that loads allowing maximum storage capacity cannot be realized with reasonable fault tolerance. In particular, the fault tolerance to initial patterns with 'false alarm' errors is very low, see Figs. 4 and 5(a). Thus, in applications, the number of stored patterns has to be drastically reduced compared with the theoretical value of M^*.

2.3. NAM models combining auto- and hetero-association

Associative memories work either hetero- or auto-associatively. In a hetero-associative memory, a mapping, $\mathscr{M}$, between pairs of different memory patterns is stored. An auto-associative memory — with the Hopfield model as the most prominent example (Hopfield, 1982) — is the special case where an identity mapping is stored, i.e. $\mathscr{S}:=\{(x^\nu,x^\nu){:}\ x^\nu\ \in\ \{0,1\}^n,\ \nu=1,\cdots,M\}$. Note that, with auto-association, (i) the synaptic memory matrix formed by Eq. (1) is symmetrical; and (ii) the retrieval Eq. (3) performs pattern completion, if noisy memory patterns are used as initial patterns. Hetero-associative retrieval provides a fault-tolerant mapping between different subrepresentations. Typically, association tasks occurring with applications are rather of the hetero-associative type.

Nevertheless, hetero-associative tasks can also be carried out in an auto-associative NAM, if the memory pairs are concatenated into larger memory patterns that are stored auto-associatively. Using an auto-associative memory allows both the mapping of initial patterns to the associated memory patterns and pattern completion on the noisy initial patterns. However, the price for this extended functionality is a larger memory matrix that has to be represented by adjustable synapses (see Fig. 1): the weights of the hetero-associative memory constitute one off-diagonal quadrant of the auto-associative matrix. Additionally, the two diagonal quadrants contain weights that store auto-associations within each subrepresentation and a second off-diagonal quadrant contains the transposed hetero-associative weights.

The combination of auto- and hetero-association in a memory only using the hetero-associative matrix can be achieved by introducing bidirectional iterative retrieval. In the bidirectional associative memory (BAM) (Kosko, 1987), two pools of neurons are connected by the hetero-associative weights in one off-diagonal quadrant of the matrix in Fig. 1 (see Section 3.1). Thus, the introduction of more sophisticated retrieval yields an extended functionality of the memory model. Moreover, as we will show, appropriate iterative retrieval strategies remedy the mentioned problems of the Willshaw model for finite system size. Different from the auto-associative

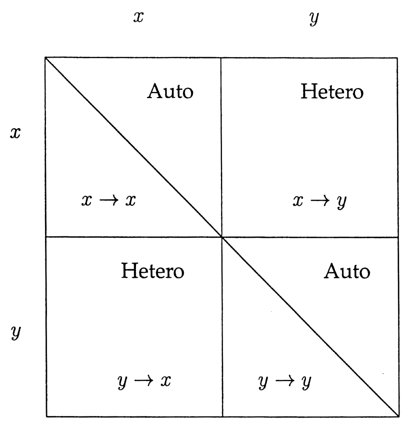

Fig. 1. A hetero-associative relation (x,y) can be stored auto-associatively, if the patterns are concatenated into larger memory patterns that are used during learning. The diagonal quadrants of the resulting memory matrix contain auto-associations: the upper left quadrant corresponds to the x patterns and the lower right quadrant to the y patterns. The off-diagonal quadrants contain hetero-associations between x and y patterns: one is the transposed matrix of the other. This associative processing provides extended functionality, namely hetero-associative mapping and auto-associative completion of x and y patterns but requires more synaptic memory. The same functionality can be achieved by BAM retrieval schemes that use only one offdiagonal quadrant of the synaptic memory.

implementation, the BAM does not store the autocorrelations within each of the neuron pools. Nevertheless, information about these auto-associations is contained in the hetero-associative weights and can be exploited by appropriate retrieval strategies (see Section 3.3).

3. Improved retrieval strategies

This section describes the model modifications that we introduce in the Willshaw model. The simplest idea is the straight-forward BAM extension which, however, turns out as an inefficient model. The literature about BAM concentrates on improving the learning prescription: either multiple training schemes have to be employed (Hassoun, 1989), or a higher number than nm weights have to be stored, for instance, if dummy augmentation (adding subsidiary components to the original memory patterns) or higher-order connections are used (Wang et al., 1990; Leung et al., 1995). Our modifications follow the other alternative: we propose more refined retrieval strategies that enhance the performance by a better employment of the information stored in the simple Hebbian second-order correlation matrix Eq. (1).

3.1. Standard bidirectional (SB) retrieval

For the hetero-associative Willshaw model, the straight-forward iterative retrieval extension is the BAM, where standard retrieval steps are performed in both directions (Haines and Hecht-Nielsen, 1988). With $x(0) = \tilde{x}$, the iterative retrieval scheme is:

$$y(r+1)_j = H[Cx(r)]_j - \Theta(r+1)\ \forall j \tag{4}$$

$$x(r+1)_i = H[C^{\mathrm{T}}y(r)]_i - \Xi(r+1)\ \forall i \tag{5}$$

With threshold choice from Section 2.2 in Eq. (4), $y(1)$ contains the memory pattern distorted only by ‘false alarm’ errors. Because of low robustness to ‘false alarm’ errors of the standard model, completion of the initial pattern and reduction of retrieval errors by bidirectional iteration is limited to the range of low memory load (see Proposition 4 in Section 4.4). With a higher memory load, iterative retrieval using a constant threshold causes either information loss or exploding activity.

3.2. Mechanisms of activity reduction

To avoid the problem of exploding activity, this sections presents two different methods to limit the activity in the network.

3.2.1. Iterative retrieval with Boolean ANDing

In a retrieval step with the ‘no misses’ threshold setting, the Willshaw model produces only ‘false alarm’ errors, cf. Section 2.2. Further iteration steps have therefore only one function: the reduction of ‘false alarm’ errors. Based on this idea, a simple trick has been used by Schwenker et al. (1996) for auto-association to prevent activity explosion. From the second iteration step onwards, Boolean ANDing with the previous pattern is applied:

$$x(r+1)_i = x(r)_i \wedge H([Cx(r)]_i - \Theta(r+1))\ \forall i \tag{6}$$

Here we use the definition $x \wedge y = xy$ for $x,y \in \{0,1\}$.

However, for bidirectional hetero-associative retrieval starting with the initial pattern $x(0)$, a previous pattern version with all correct one components present is not available before the third retrieval step ($r = 3$). Therefore, in the iteration scheme Eqs. (4) and (5), Boolean ANDing can only be used from the third iteration step onwards, which may be too late to prevent activity explosion if the memory load is high.

3.2.2. Pattern part retrieval

Alternatively, one can restrict the activity obtained by the first step, let us say to the value $f < b$: the corresponding retrieval task is to determine $f < b$ components that agree with the highest probability with 1-elements in the y-memory pattern using one-step retrieval. In the case of a perfect hit, the extracted elements are a part of the memory pattern with size f, and therefore we call this task pattern part retrieval. If the number of ‘false alarm’ and ‘miss’ errors in the initial pattern is denoted by g and z, respectively, there are two different regimes: for $g = 0$ the highest possible threshold is the ‘no misses’ threshold setting: $\Theta = a - z$, cf. Section 2.2, leading to a retrieved pattern with activity larger than or equal to b. From this pattern, one has

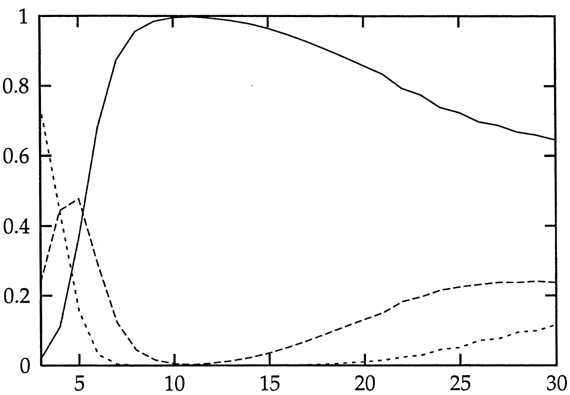

Fig. 2. Hit distribution of pattern part retrieval extracting exactly f elements that belong with highest probability to the memory pattern as described in Section 3.2.2. Proposition 4.3 is applied with $f = 2$, and the parameters $n = m = 2000$, $a = b = 10$ and $M = 15\,000$. The x-axis displays the activity of the initial pattern. Activities lower than $a = 10$ correspond to initial pattern with no 'add' errors, i.e. $g = 0$, activities higher than $a = 10$ correspond to 'miss' error free initial patterns, i.e. $z = 0$. The dotted curve displays $P_S(g,z,2;0)$, the dashed curve $P_S(g,z,2;1)$ and the solid curve $P_S(g,z,2;2)$.

to select randomly the f elements. For $g > 0$, there is more freedom in the possible threshold choice; one-step retrieval with $a - z > \Theta \geq a - z + g$ yields retrieval patterns with reduced activity. Reducing the activity by increasing the threshold yields a higher hit probability compared with random selection of the f elements in the 'no misses' retrieval pattern, see results of Proposition 4.3 in Fig. 2.

3.3. Conditional links

The Willshaw model classifies an output neuron as active depending on the dendritic sum Eq. (2), i.e. the overlap between the corresponding matrix column and the input pattern. The new retrieval methods that will be proposed in the next section employ a more selective classification evaluating overlaps between parts of different columns of the clipped Hebbian matrix: for a given $x(r)$ we call the underlying quantities conditional links:

$$w_{jl}^{x(r)} = \sum_i C_{ij} C_{il} x(r)_i \ \forall j, \ \ l \in \{1, \ldots, m\} \tag{7}$$

The conditional link $w_{jl}^{x(r)}$ is a nonnegative integer and expresses the overlap between the matrix columns j and l conditioned on the pattern $x(r)$, that is, restricted to such matrix rows i where $x(r)_i = 1$. Conditional links can be viewed as virtual lateral interactions between units in the y-pattern. Since $|w_{jl}^{x(r)}| \leq |x(r)|$ and $|w_{jl}^{x(r)}| = |x(r)|$, if $[x(r) = x^\nu] \cap [y_j^\nu = y_l^\nu = 1]$, it is clear that a high conditional link indicates a high probability that both components in the y pattern belong to one learning pattern y^ν, under the condition that $x(r)$ has a high overlap with the corresponding learning pattern x^ν. The diagonal interactions are equal to the standard dendritic sums caused by $x(r)$, i.e. $w_{jj}^{x(r)} = [Cx(r)]_j$ as in Eq. (4).

Of course, for a bidirectional retrieval scheme, conditional links can similarly be computed in the x-layer for a given pattern $y(r)$:

$$w_{ik}^{y(r)} = \sum_j C_{ij} C_{kj} y(r)_j \ \forall i, \ \ k \in \{1, \ldots, n\} \tag{8}$$

The detection of active components in a pattern can now be achieved by finding the clique of neurons that are tied together by the highest conditional links. This problem can be approximately solved by different bidirectional retrieval schemes involving computation of weighted sums with matrix rows and columns. For this reason, retrieval with conditional links is referred to as crosswise bidirectional (CB) retrieval and its different variants will be explained in the next section.

3.4. Crosswise bidirectional (CB) retrieval

Crosswise bidirectional (CB) retrieval estimates a solution of the conditional links clique problem by auto-associative retrieval using the virtual feedback network with interaction matrix w and threshold neurons.

The dendritic sum of CB retrieval can be expressed as:

$$[w^{y(r+1)} x(r)]_i = \sum_k w_{ik}^{y(r+1)} x_k = \sum_{j \in y(r+1)} C_{ij} [Cx(r)]_j \tag{9}$$

In Eq. (9), $[Cx(r)]_i$ is the dendritic sum of the Willshaw model Eq. (2) that is propagated back through the synaptic weights. It has to be emphasized that Eq. (9) is not just the linear transformation $C^T{*}C$: the signal flow is gated by $y(r + 1)$ allowing only feedback of dendritic sums at active neurons in this pattern. Thus, the RHS of Eq. (9) gives a parallel scheme for fast computation of the dendritic sums during CB retrieval.

There are two CB retrieval schemes how a fixed point pattern pair can be approached either by a sequence of auto-associative fixed point retrieval processes (variant I) or by direct hetero-associative iteration to the fixed point (variant II). Both variants have been tested experimentally, see Section 5. In the dendritic sum Eq. (9) of CB retrieval an integer valued vector $[Cx(r)]_j\, y_j(r + 1)$ has to be propagated through the synaptic weights.

3.4.1. Variant I

The initial pattern $x(0)$ is used in a pattern part retrieval step, cf. Section 3.2.2, to estimate one component of the y pattern. The resulting pattern pair is put as $[x(0),y(-1)]$ in Eq. (10) of the following iterative scheme to obtain $y(1)$.

Using the pattern pair $[x(r),y(r - 1)]$, the pattern $y(r + 1)$ is the fixed point after iteration of:

$$y(s + 1)_j = H\Big\{[w^{x(r)} y(s)]_j - \Theta(s + 1)\Big\} \ \forall j \tag{10}$$

starting with $y(0) = y(r - 1)$.

Using the pattern pair $[y(r),x(r - 1)]$, the pattern $x(r + 1)$

is the fixed point after iteration of:

$$x(s+1)_i = H\Big\{[w^{y(r)}x(s)]_i - \Xi(s+1)\Big\}\ \forall i \qquad (11)$$

starting with $x(0) = x(r-1)$.

3.4.1.1. Threshold strategy. For pattern part retrieval the highest possible threshold is chosen that yields nonzero activity. In Eqs. (10) and (11), fixed points are traced at threshold values $b|x(r)|$ and $a|y(r)|$, respectively.

3.4.2. Variant II

This variant includes also Boolean ANDing from Section 3.2.1. Starting again from $x(0)$ with a standard retrieval step Eq. (4), the resulting pattern pair is used as $[x(0),y(1)]$ in a single step of Eq. (11) to determine $x(2)$. For $r > 2$, the update is performed by the iterative retrieval equations:

$$y(r+2)_j = y(r)_j \wedge H\Big\{[w^{x(r+1)}y(r)]_j - \Theta(r+2)\Big\}\ \forall j \qquad (12)$$

$$x(r+2)_i = x(r)_i \wedge H\Big\{[w^{y(r+1)}x(r)]_i - \Xi(r+2)\Big\}\ \forall i \qquad (13)$$

Unlike in variant I in Eqs. (12) and (13), only one auto-associative step is performed, and no iteration takes place until the fixed point, as in variant I.

3.4.2.1. Threshold strategy. For $r = 1$, a standard retrieval step is performed with threshold as chosen in Section 2.2, yielding a pattern with $|y(1) \geqslant b$. For initial patterns with $|x(0)| < a$, the $r = 2$ update step Eq. (13) is carried out without Boolean ANDing. In this iteration step, the threshold $\Xi(2) = |x(0)||y(1)|$ is used, yielding a result with $|x(2)| \geqslant |x(0)|$.

Subsequently, and in the case of initial patterns with $|x(0)| \geq a$ already for $r = 2$, the thresholds $\Xi(r)$ and $\Theta(r)$ are adjusted in order to obtain resulting activities which are as close as possible to prescribed values. During the iteration, these values are chosen as follows:

$$\Xi(r+1) : |x(r+1)| \simeq |x(r)| - 1$$

$$\Theta(r+1) : |y(r+1)| \simeq b$$

The iteration is stopped if $|x(r)| = a$ is reached.

4. Model analysis

4.1. Definitions

In the following, we assume that the memory patterns are randomly generated, i.e. each component in the x- and y-patterns has been set to 'one' with probability $p: = a/n$ and $q: = b/m$ respectively. (a, b, m, and n as defined in Section 2.2).

We consider the retrieval with an initial pattern $x(0) = \tilde{x}^{\mu}$ which is the noisy version of the x pattern in a learning pattern pair (x^{μ}, y^{μ}). For the occurrence probabilities of 'false alarm' and 'miss' errors, the following notation will be used:For r odd:

$$\alpha(r) := p[y(r)_j = 1|y_j^{\mu} = 0]$$

$$\beta(r) := p[y(r)_j = 0|y_j^{\mu} = 1]$$

For r even:

$$\gamma(r) := p[x(r)_j = 1|x_j^{\mu} = 0]$$

$$\zeta(r) := p[x(r)_j = 0|x_j^{\mu} = 1]$$

In this notation, $r = 0$ plays a unique role: $\gamma^{\mathrm{I}}: = \gamma(0)$ and $\zeta^{\mathrm{I}}: = \zeta(0)$ describe the initial errors, i.e. the error levels in the initial pattern $x(0)$. It has to be kept in mind that the error probabilities at time r depend on the initial errors as well as on the threshold values. For instance, for odd r's we have:

$$\alpha(r) = f[\Theta(r), \Theta(r-2), \cdots, \Theta(1), \Xi(r-1), \Xi(r-3), \cdots, \Xi(2), \gamma^{\mathrm{I}}, \zeta^{\mathrm{I}}] \qquad (14)$$

A NAM model works efficiently, if the set of retrieved patterns contains as much information as possible about the memories, and as few as possible synaptic memory have been used: the information efficiency of a memory model is defined as the dimensionless ratio between the information contained in the retrieved patterns and in the synaptic memory matrix, respectively. The information about a memory pattern that is contained in a noisy version of this pattern can be calculated for each component in terms of the transinformation $t(p,\alpha,\beta): = i(p) - i(p,\alpha,\beta)$. Here, $i(p): = -p\log_2[p] - (1-p)\log_2[1-p]$ is the Shannon information in a pattern component (Shannon and Weaver, 1949). The conditional information describes the amount of information necessary to correct the errors:

$$i(p,\alpha,\beta) := \tilde{p}i\left(\frac{(1-p)\alpha}{\tilde{p}}\right) + (1-\tilde{p})i\left(\frac{p\beta}{1-\tilde{p}}\right) \qquad (15)$$

with $\tilde{p} := p(1-\beta) + (1-p)\alpha$.

Depending on the considered memory task, the net information provided by the retrieval about the memory patterns has to account for the effect of retrieval errors, and the information already contained in the initial patterns. Usually, this net information divided by the weight matrix size nm is called memory capacity and used as a performance measure for NAM. The information efficiency is then simply the memory capacity divided by the number of bits needed to specify a synaptic value — and, of course, for NAM with binary synapses, both quantities coincide. For a hetero-associative mapping task, we define the output capacity as:

$$A := Mm\ t(q, \alpha(r), \beta(r))/mn \qquad (16)$$

for large $r(r \to \infty)$. The maximum output capacity will be expected, if no fault tolerance has to be provided, that is, with noiseless initial patterns. For an auto-associative completion task acting on distorted x-patterns, the completion capacity is the information balance defined by:

$$C := Mn[t(p,\gamma(r),\zeta(r)) - t(p,\gamma^{\mathrm{I}},\zeta^{\mathrm{I}})]/n^2 \tag{17}$$

for large $r(r \to \infty)$. For the BAM situation, when the memory should complete the input pattern and, at the same time, map to the output pattern we define the search capacity: $S := C + A$. The search capacity that equals the information efficiency for BAM models with binary synapses will be used to evaluate and compare the different models considered in this paper.

4.2. Asymptotic information efficiency of the Willshaw BAM

Former theoretical results can be used to derive asymptotic values for the information efficiency of BAM: (a) for the case of error-free address, that is, $\gamma^{\mathrm{I}} = 0$ and $\zeta^{\mathrm{I}} = 0$, since $C(0,0) = 0$, the search capacity is limited by the Willshaw capacity: $S(0,0) = A(0,0) = \ln[2]$ bit/synapse (Willshaw et al., 1969); (b) addresses with $\gamma^{\mathrm{I}} = 0$ and $\zeta^{\mathrm{I}} = 0.5$ achieve asymptotically the maximum completion capacity with an asymptotic value of $C(0,0.5) \leq \ln[2]/4$ bit/synapse. With the same input noise the asymptotic output capacity is given by $A(0,0.5) \leq \ln[2]/2$ bit/synapse (Palm, 1988). Thus, the asymptotic search capacity for bidirectional retrieval is $S(0,0.5) = (3\ln[2])/4 = 0.52$ bit/synapse.

A universal upper bound on the search capacity can be obtained by analyzing a process where the memory recognizes stored patterns in the whole space of sparse initial patterns: an initial pattern is classified as known if it is reproduced after a bidirectional retrieval cycle. The information capacity of this recognition process is an upper bound of the completion capacity, and has been determined as ln[2]/2 (Palm and Sommer, 1992), which is achieved as well with the parameter set M,p,q providing $A = \ln[2]/2$. Thus, the asymptotic search capacity $S(0,0.5)$ is bounded by ln[2] bit/synapse. Indeed, it can be shown that this bound cannot be exceeded by any choice of parameter values. Again, the Willshaw capacity (Willshaw et al., 1969) turns out as an invincible bound for retrieval from the binary-valued synaptic weights. From the finite auto-associative model, we know that iterative retrieval methods are able to reach and even slightly surpass the capacity values calculated for the infinite model (Schwenker et al., 1996). In fact, only iterative retrieval provides the retrieval error reduction necessary to exploit the high capacity of the model. The error probabilities of one-step retrieval become negligible only for system sizes far beyond realization.

4.3. Refined combinatorial analysis for the Willshaw model

For a retrieval error analysis of the Willshaw model, we have to consider the rows in each synaptic matrix column, which correspond to the 'one' components in the initial pattern $x(0)$. If $c = |x(0)|$, one has to calculate the probability that a c-subcolumn contains d 'one' entries $P(c;d) := \mathrm{p}[\sum_{i\in x(0)} C_{ij} = d]$. For a matrix where the elements are generated independently with $p_1 := \mathrm{p}[C_{ij} = 1]\forall i,j$, the integer d is binomially distributed with $P_{\mathrm{B}}(p_1,c;d)$.

The binomial distribution is defined as

$$P_{\mathrm{B}}(p,c;d) = \binom{c}{d} p^d (1-p)^{c-d},$$

where

$$\binom{c}{d} = \binom{c}{d, c-d}$$

denotes the binomial coefficient, a special case of the multinomial coefficient that will be used later:

$$\binom{c}{d_1,\cdots,d_k} = \frac{n!}{d_1!d_2!\cdots d_k!}\ \forall\{d_1,\cdots,d_k : \sum_{i=1}^{k} d_i = c\}.$$

For discrete distributions we use consistently the notation $P_{\mathrm{D}}(x_1,\cdots,x_n,c;d)$: the variable d can assume the values $0,\cdots,c$, $x_1,\cdots,x_n$ are additional parameters, and D is a descriptor of the distribution, if it is not uniquely specified by the number of parameters, for instance, subscript B stands for the binomial distribution, and S for the distribution of the dendritic sum in standard retrieval. For the cumulation[3] of a discrete distribution, we will use a similar nomenclature: $Q_{\mathrm{D}}(x_1,\cdots,x_n,c;d) = \sum_{i=d}^{c} P_{\mathrm{D}}(x_1,\cdots,x_n,c;d)$. For the sake of brevity, dependencies on M, n, m, p, q will be suppressed in the parameter lists.

In the first analysis of the Willshaw model (Willshaw et al., 1969), the binomial distribution has been used with the estimation:

$$p_1 \simeq 1 - (1-pq)^M \tag{18}$$

Since, even with random patterns, Eq. (1) does not independently generate the matrix elements, the analysis should be refined. Proposition 4.1 gives the distribution of 1-elements in a column (or row) in the synaptic weight matrix after storage of random patterns. It is a prerequisite for the three subsequent propositions: Proposition 4.2 derives the retrieval errors in the first retrieval step, i.e. $r = 1$. All iterative retrieval methods we have considered deviate only from the second retrieval step onwards. Proposition 4.3 analyzes the pattern part retrieval and Proposition 4.4 yields lower bounds for the retrieval errors with SB retrieval.

Proposition 4.1. (Willshaw model: distribution of dendritic sum.) *After learning according to Eq. (1), the probability that a synaptic c-subcolumn contains exactly $d \leq c$ one entries is:*

$$P_{\mathrm{S}}(c;d) = \binom{c}{d} \sum_{s=0}^{d} (-1)^s \binom{d}{s} \left[1 - q\left\{1-(1-p)^{s+c-d}\right\}\right]^M \tag{19}$$

[3] For convenience, we do not use the standard definition of cumulation: with our definition $Q(c;d) = 1\ \forall\ d \leq 0$, whereas in the usual definition the sum runs over the lower interval, i.e. $Q'(c;d) = 1\ \forall\ d \geq c$.

Proof. Eq. (19) can be derived from the formula given by Buckingham and Willshaw (1992), rewritten in the following as Eq. (20). The dependencies between matrix elements caused by the clipped Hebbian learning process are taken into account by introducing the unit usage, i.e. the number of memory patterns that have changed the matrix column of a particular neural unit. For random patterns, the unit usage is a binomially distributed quantity. In Eq. (20), the 1-density in a synaptic subcolumn for a fixed value of u is averaged over all its possible values:

$$\frac{P_S(c;d)}{\binom{c}{d}} = \sum_{u=1}^{M} P_{\mathrm{B}}(q,M;u) \underbrace{[1-(1-p)^u]^d}_{=\sum_{s=0}^{d}\binom{d}{s}(-1)^s(1-p)^{us}} [(1-p)^u]^{c-d}$$

$$= \sum_{s=0}^{d}\binom{d}{s}(-1)^s(1-q)^M \times \underbrace{\sum_{u=1}^{M}\binom{M}{u}\left[\frac{q}{1-q}(1-p)^{c-d+s}\right]^u}_{=\left[1+\frac{q}{1-q}(1-p)^{c-d+s}\right]^M-1} \qquad (20)$$

The derivation of Eq. (19) from Eq. (20) employs a twofold application of the binomial theorem as highlighted by the underbracings.

Proposition 4.1 provides a big difference in numerical evaluation time for the parameter range where the Willshaw model has high information capacity: in Eq. (20), M terms have to be added with typically $M \propto n^2/(\ln[n])^2$, while in Eq. (19), the sum is only conducted over d terms with typically $d \propto \ln[n]$ (for the determination of typical values, see Palm and Sommer, 1995).

Proposition 4.2. (Willshaw model: retrieval errors.) *If the initial pattern contains g 'false alarm' and z 'miss' errors and the threshold is set to* $\Theta(1) = \Theta$*, the retrieval error probabilities are:*

$$\alpha_{\mathrm{S}}(g,z,\Theta) = Q_{\mathrm{S}}(a+g-z;\Theta) \qquad (21)$$

$$\beta_{\mathrm{S}}(g,z,\Theta) = 1 - Q_{\mathrm{S}}(g;\Theta-(a-z)) \qquad (22)$$

with the cumulative distribution:

$$Q_{\mathrm{S}}(c;\Theta) = 1 + \binom{c}{\Theta}\sum_{i=0}^{\Theta-1}(-1)^{\Theta-i} \times \binom{\Theta}{i}\frac{\Theta-i}{c-i}\left[1-q\left\{1-(1-p)^{c-i}\right\}\right]^M \qquad (23)$$

Proof. The distribution of the dendritic sum of standard retrieval Eq. (19) given in Proposition 4.1 has to be cumulated for all values larger than the threshold to calculate the 'false alarm' error probability Eq. (21). The terms in the resulting double sum can be relabeled in order to avoid multiple terms with the same factor $[1-q\{1-(1-p)^j\}]^M$. Beyond the basic relations for binomial coefficients, the key transformations to obtain the cumulative distribution Eq. (23) are:

$$\sum_{v=0}^{w}\binom{w}{v}(-1)^v = \delta_{w,0} \quad \sum_{v=0}^{w-t}\binom{w}{v}(-1)^{w-t-v} = \binom{w-1}{t-1} \qquad (24)$$

with $\delta_{w,v}$ the Kronecker symbol. An invaluable source for combinatorial identities is Riordan (1968): relations of Eq. (24) can be found on pages 4 and 34. In our calculation, they are applied for $w = c - i$ and $t = \Theta - i$. The 'miss' error probability also requires cumulation of Eq. (19) of Proposition 4.1 and can be computed similarly using Eq. (23).

In the parameter range considered in our simulation experiments, the numerical evaluation at a reasonable speed is only possible using Propositions 4.1 and 4.2. Eq. (21) has been compared with the formula obtained by cumulating the distribution [Eq. (20)] in terms of evaluation time. In addition to the shorter sum in Eq. (19), in (21) the double sum is reduced to a summation over a single variable. With Mathematica,[4] the evaluation time for a single value is reduced from minutes to the fraction of a second. Eq. (19) generalizes an old result that had been derived for the special case $c = d$ (and exactly equal activities of the memories) by Palm (1980). We will now use Propositions 4.1 and 4.2 to analyze the pattern part retrieval.

Proposition 4.3. (Willshaw model: hit distribution of pattern part retrieval.) *Employing pattern part retrieval, described in Section 3.2.2 with g 'false alarm' and z 'miss' initial errors, the probability that a pattern part with size* $f < b$ *contains* $e \le f$ *elements that belong to the learning pattern is:*

$$P_S(g,z,f;e) = \begin{cases} \dfrac{P_{\mathrm{H}}(m,b,f;e)}{g}\displaystyle\sum_{i=1}^{g}\left(\frac{\mathscr{R}(g,z,a-z+i)^e}{\sum_{j=0}^{f}P_{\mathrm{H}}(m,b,f;j)\mathscr{R}(g,z,a-z+i)^j}\right) & g>0 \\ P_{\mathrm{B}}(p_r(z),f;e) & g=0 \end{cases} \qquad (25)$$

[4] To evaluate the binomial sums it is important to use a programming language where the computation precision can be increased arbitrarily.

with $\mathscr{R}(g,z,\Theta) = [1 - \beta_S(g,z,\Theta)][1 - \alpha_S(g,z,\Theta)]/[\alpha_S(g,z,\Theta)\beta_S(g,z,\Theta)]$, $p_r(z) = q/[q + \alpha_S(0,z,a-z)]$, $\alpha_S(g,z,\Theta)$ *and* $\beta_S(g,z,\Theta)$ *from Proposition 4.2, and with the hypergeometrical distribution*

$$P_H(m,b,f;e) = \binom{b}{e}\binom{m-b}{f-e}/\binom{m}{f}.$$

Proof. The wanted distribution can be calculated as conditional probability: $P_S(g,z,f;e) = p[e|f]$. For $g > 0$, it is $p[e|f] = \sum_\Theta p[e,f,\Theta]p[\Theta]/p[f,\Theta]$, with $p[e,f,\Theta] = p[e|\Theta]p[w = f - e|\Theta]p[\Theta]$ and $p[f,\Theta] = p[f|\Theta]p[\Theta]$. The conditional distribution of correct ones is given by $p[e|\Theta] = P_B[1 - \beta_S(g,z,\Theta),b;e]$, the conditional distribution of wrong ones by $p[w|\Theta] = P_B[\alpha_S(g,z,\Theta),m - b;w]$, and $p[f|\Theta] = \sum_{e=0}^{f} p[e|\Theta]p[w = f - e|\Theta]$. For $g = 0$, one-step retrieval is performed with the highest possible threshold: $\Theta = a - z$. From the resulting pattern, the f one components have to be chosen randomly, leading to a hit probability for a single component of $p_r(z)$.

4.4. Error analysis of SB retrieval

Proposition 4.4. (SB retrieval errors: lower bounds.) *The 'false alarm' error probabilities after the first step, i.e.* $r \geq 1$ *satisfy:*

$$\gamma(r) \geq Q_S(|y(r-1)|; \Xi(r))\ \forall r \text{ even} \tag{26}$$

$$\alpha(r) \geq Q_S(|x(r-1)|; \Theta(r))\ \forall r \text{ odd} \tag{27}$$

with the definitions of $Q_S(c;\Theta)$ *from Proposition 4.2.*

Proof. In the second step of SB retrieval, we have to consider two different cases in the update process: all neurons not belonging to the set $x^\mu \cup x(0)$ can be described with Eq. (19) as in the first retrieval step, since all synapses ending at such neurons have not been involved to obtain $y(1)$. Neurons in the set $x^\mu \cup x(0)$ will behave differently because of the statistical dependencies between $y(1)$ and the matrix elements. Since the synapses corresponding to this set have been selected by the threshold criterion during the first retrieval step, the 'false alarm' error probability of these units will be strictly increased. Thus, the cumulative distribution from Proposition 4.2 provides a lower bound on the error probability in the second step. This argument for the second step can be extended to subsequent iteration steps. Hence, the one-step error is always a lower error bound.

As pointed out in Section 2.2, the most urgent demand on modified retrieval strategies is the improved ability to process initial patterns with 'false alarm' errors. We now consider initial patterns without 'miss' errors, i.e. $\zeta^I = 0$. For this case and with our assumption that the x and y patterns have the same dimension and activity, bidirectional iteration of standard retrieval is useful, if the noise is decreasing during iteration, i.e.:

$$\gamma^I > \alpha(1) > \gamma(2) > \cdots \tag{28}$$

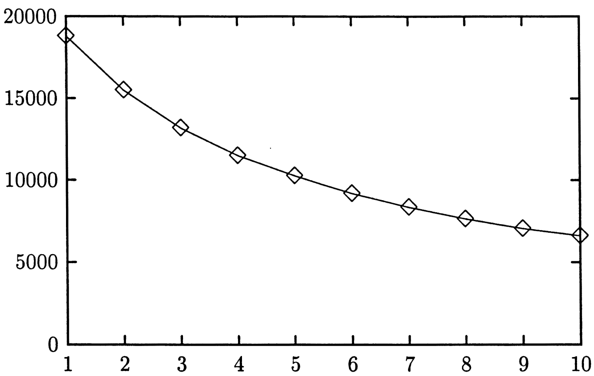

Fig. 3. Maximum number of patterns M_{max} for which bidirectional iteration of standard retrieval can be expected to improve the retrieval results. In this calculation, we use again the parameters $n = m = 2000$ and $a = b = 10$. The x-axis displays the number g of 'false alarm' errors in the initial pattern. The initial 'miss' errors are set to zero, i.e. $z = 0$, thus the initial activities vary in the range $11 \leq |x(0)| \leq 20$.

For a given initial 'add' noise level, $\gamma^I = g/(n - a)$, a necessary condition for decreasing noise can be derived from Proposition 4.3, yielding the fixed point equation:

$$g = (n-a)Q_S(a+g;a) \tag{29}$$

where $Q_S(c;\Theta)$ is given by Proposition 4.2. Solving Eq. (29) for each number of initial errors, $g > 0$ gives an upper bound on the number of possible stored patterns with standard bidirectional retrieval. Beyond that bound, which is displayed in Fig. 3, an improvement by iteration of retrieval steps has to be excluded. As a consequence, the number of stored patterns has to stay below that curve. Since the bound decreases rapidly with a growing number of initial errors, standard bidirectional retrieval cannot improve the memory performance.

4.5. Error analysis of CB retrieval

Since the analysis CB retrieval is more intricate, the proposition in this section will neglect the mutual dependencies between different elements in the Hebbian matrix.

Proposition 4.5. (CB retrieval: errors after first step of variant I.) *The initial pattern contains g 'false alarm' and z 'miss' errors. Consider the first step of variant I, i.e. Eq. (10) for s = 0. Assume that y(0) computed by pattern part retrieval has activity* $f \geq 1$ *and contains* $e \leq f$ *correct 1-elements. The error probabilities of the first update step of CB retrieval are given by:*

$$\alpha_{CB}(g,z,f,e,\Theta) = Q^{\alpha}_{CB}(e,(a-z+g)f;\Theta) \tag{30}$$

$$\beta_{CB}(g,z,f,e,\Theta) = 1 - Q^{\beta}_{CB}(e,(a-z+g)f;\Theta) \tag{31}$$

where the distributions of the dendritic potentials that have *to be cumulated are:*

$$P^{\rho}_{\mathrm{CB}}(e,(a-z+g)f;\Theta) = \sum_{i=\max[0,\Theta-gf]}^{\min[(a-z)f,\Theta]} P^{\rho}_{\mathrm{SRI}}(e,(a-z)f;i)P_{\mathrm{SRI}}(gf;\Theta-i) \quad (32)$$

with $\rho=\alpha,\beta$. *The first factor in the convolution sum Eq. (32) is:*

$$P^{\alpha}_{\mathrm{SRI}}(e,df;\Theta) = \left(\frac{p_1}{1-p_1}\right)^{\Theta}\left[\frac{(1-p_1)^f}{p_1^{e-1}}\right]^d Z^{\alpha}[f-e,f,e,\Theta,d] \quad (33)$$

$$P^{\beta}_{\mathrm{SRI}}(e,df;\Theta) = \left(\frac{p_1}{1-p_1}\right)^{\Theta}\left[\frac{(1-p_1)^f}{p_1^{e}}\right]^d Z^{\beta}[f-e,f,e,\Theta,d] \quad (34)$$

where Z^{α} *and* Z^{β} *can be calculated recursively:*

$$Z^{\rho}[f,f^*,e,\Theta,d] = \sum_{i=0}^{Fl[\Theta/f]} \frac{\binom{f^*-e}{f}^i}{i!} Z^{\rho}[f-1,f^*,e,\Theta-(f+e)i,d-i]$$

with initial values:

$$Z^{\alpha}[0,f^*,e,\Theta,d] = H\left[d-\frac{\Theta}{e}\right]\delta_{\frac{\Theta}{e},Fl\left[\frac{\Theta}{e}\right]}\binom{d}{\Theta/e} \times \left[\frac{p_1^{e-1}}{(1-p_1)^{f^*-1}}\right]^{d-\Theta/e}$$

$$Z^{\beta}[0,f^*,e,\Theta,d] = \delta_{\frac{\Theta}{e},d}\binom{d}{\Theta/e}\frac{1}{(\Theta/d)!}$$

Similarly, the second factor in Eq. (32) can be calculated recursively:

$$P_{\mathrm{SRI}}(df;\Theta) = \left(\frac{p_1}{1-p_1}\right)^{\Theta}\left[p_1(1-p_1)^f\right]^d Z[f,f,\Theta,d]$$

$$Z[f,f^*,\Theta,d] = \sum_{i=0}^{Fl[\Theta/f]} \frac{\binom{f^*}{f}^i}{i!} Z[f-1,f^*,\Theta-fi,d-i]$$

$$Z[1,f^*,\Theta,d] = H[d-\Theta]\binom{d}{\Theta}(f^*)^{\Theta} \times \left[1+\frac{1}{p_1(1-p_1)^{f^*-1}}\right]^{d-\Theta} \quad (35)$$

where p_1 *is the density of one entry in matrix Eq. (18), and Fl[x] is the greatest integer less or equal to x.*

Proof. The dendritic sum at each neuron is composed of two contributions, r_1 and r_2, corresponding to addressed matrix rows either agreeing with one components in the x-memory pattern or not. Each contribution is a sum over discrete random variables. For r_1, the sum contains $d=a-z$ random integers that have at on-neurons the distribution:

$$P^{\beta}_{\mathrm{RI}}(e,f;j) = P_{\mathrm{B}}(p_1,f;j-e) \quad (36)$$

and at off-neurons the distribution:

$$P^{\alpha}_{\mathrm{RI}}(e,f;j) = \begin{cases} p_1P_{\mathrm{B}}(p_1,f;j-e) & e\le j\le f \\ 1-p_1 & j=0 \end{cases} \quad (37)$$

For r_2, the sum contains $d=g$ random integers with distribution:

$$P_{\mathrm{RI}}(f;j) = \begin{cases} p_1P_{\mathrm{B}}(p_1,f;j) & 0<j\le f \\ 1-p_1[1-P_{\mathrm{B}}(p_1,f;0) & j=0 \end{cases} \quad (38)$$

A sum $r=\sum_{i=1}^d v_i$ over independent random integers $v_i=\{0,1,\cdots,f\}$ has the distribution:

$$P^{(\rho)}_{\mathrm{SRI}}(df;j) = \sum_{u\in U_d}\delta_{\Theta,\sum_j ju_j}\binom{d}{u_0,\cdots,u_l}\prod_{j=0}^{l}[P^{(\rho)}_{\mathrm{RI}}(f;j)]^{u_j} \quad (39)$$

where $U_d=\{u\in\mathbb{N}^{f+1}:\sum_{j=0}^f u_j=d\}$ is the set of vectors generating all possible configurations of the sum by $r=\sum_{j=0}^f ju_j$, and the multinomial coefficient $\binom{d}{u_0,\cdots,u_l}$ as defined in Section 4.3. The recursive formulae Eqs. (33)–(35) are obtained by insertion of Eqs. (36)–(38) in Eq. (39) and considerable algebra. The distribution of the sum of r_1 and r_2 is given by the convolution of their distributions leading to Eq. (30).

Finally, using the results of Propositions 4.3 and 4.5, we can estimate the error probabilities of the first update step of variant I, i.e. Eq. (10) with $s=0$, with the pattern $y(0)$ generated by pattern part retrieval. Corresponding to the ‘no misses’ threshold setting in simple retrieval we consider a ‘minimal misses’ threshold setting for CB retrieval that avoids miss errors whenever possible, namely in all cases where $y(0)$ contains any hits: $e>0$. For $e=0$ we assume the worst case, i.e. $\alpha_{\mathrm{CB}}=\beta_{\mathrm{CB}}=1$. The error probabilities are then:

$$\alpha_{\mathrm{CB}}(g,z,f) = P_{\mathrm{S}}(g,z,f;0)+\sum_{e=1}^{f}P_{\mathrm{S}}(g,z,f;e)\alpha_{\mathrm{CB}}(g,z,f,e,(a-z)e) \quad (40)$$

$$\beta_{\mathrm{CB}}(g,z,f) = P_{\mathrm{S}}(g,z,f;0) \quad (41)$$

Fig. 4 shows the improvement of CB retrieval in three stages: after the first update step; after the first auto-associative cycle; and after the complete procedure. As

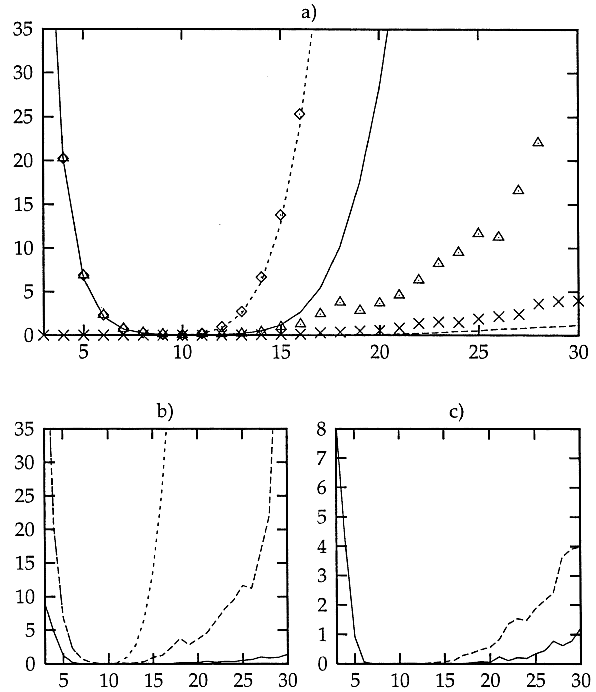

Fig. 4. Comparison of the expected number of retrieval errors in the y-patterns for the different methods. The parameters n,m,a,b,M and the x-axes labeling by the initial activity has been chosen as in Fig. 2. Points in diagram (a) display experimental errors for standard retrieval (◇) and the first auto-associative cycle of CB retrieval (△ and × for add and miss errors, respectively). Standard retrieval uses 'no misses' threshold setting and CB retrieval variant I was performed to the fixed point of the first auto-associative iteration of Eq. (10). The starting y-pattern of CB retrieval was determined by pattern-part retrieval, with $f = 3$, cf. Section 3.2.2. The lines in diagram (a) correspond to theoretical values, the dotted line displays add errors for standard retrieval [Eq. (21)], the solid and the dashed lines represent add and miss errors, respectively, produced by the first CB update step using the 'minimal misses' strategy explained in Section 4.5, and calculated by Eq. (40) and Eq. (41). The theory confirms that already the first update step outperforms standard retrieval. Note that the results after the first cycle are even better. Diagram (b) and (c) compare add and miss errors, respectively, for the different retrieval strategies. The dotted line represents results with standard retrieval, dashed lines with the first iteration cycle of CB retrieval and solid lines with complete CB retrieval. Clearly the fully iterated (complete) CB retrieval yields the lowest retrieval errors.

can be observed in Fig. 4(a), the theory predicts lower errors of the first CB update step compared with standard retrieval. Experimental results indicate that errors of the first CB retrieval cycle are below these of a single CB update step, and fully iterated CB retrieval displayed in diagrams (b) and (c) yield by far the lowest retrieval errors.

4.6. The problem of optimal retrieval (POOR)

Since all bidirectional retrieval strategies start with an ordinary threshold retrieval step, as analyzed by Proposition 4.2, condition Eq. (28) is a universal prerequisite for successful retrieval for initial patterns without miss errors. If $b = a$ and $m = n$, it should not matter whether you start with an x or a y initial pattern and for error-free initial patterns, i.e. $\gamma^{\mathrm{I}} = 0$ or $\zeta^{\mathrm{I}} = 0$, requirement Eq. (28) demands that learning pattern pairs are fixed points of the synaptic matrix:

$$y^* = H([Cx^*] - \Theta);\ x^* = H([C^T y^*] - \Xi) \qquad (42)$$

We call (x^*,y^*) in Eq. (42) a (Θ,Ξ)-fixed point of the synaptic matrix C. $(|x^*|,|y^*|)$-fixed points we call simply the fixed points of the synaptic matrix C, denoted by $\mathscr{F}$. A necessary condition, that the associative memory at least 'recognizes' all learned patterns, is:

$$\mathscr{S} \subseteq \mathscr{F} \qquad (43)$$

On the other hand, no pattern different from the learning patterns, often denoted as spurious state, should fulfill the fixed point condition:

$$\mathscr{F} \subseteq \mathscr{S} \qquad (44)$$

If condition Eq. (44) is violated, the set of initial patterns from which the retrieval dynamics cannot move to the nearest learning pattern is nonempty. Combining both conditions, Eq. (43) and (44), yields:

$$\mathscr{F} = \mathscr{S} \qquad (45)$$

which is a necessary condition that individual learning patterns can be retrieved in an effective and fault-tolerant way from the synaptic matrix. The Willshaw matrix fulfills Eq. (45), if the memory load is below the point where superpositions of traces form completely filled 'subrectangles'.

Given that the matrix fulfills Eq. (45), what can be optimally expected from a retrieval procedure? The answer can be formulated as the following linear optimization problem:

> Problem of optimal retrieval (POOR): given a binary matrix, C_{ij}, with the fixed points set, $\mathscr{S}$, and an initial pattern, $x(0)$, find a pattern pair (x^*,y^*) with:
>
> $$\langle x^*, x(0)\rangle = \text{Max!} \qquad (46)$$
>
> which satisfies the constraints: $|x^*| = a$, $|y^*| = b$, and (x^*,y^*) is a fixed point of C.

In POOR, the maximization guarantees that x^* is the closest fixed point, i.e. with maximum overlap to the initial pattern $x(0)$. The normalizations and the fixed point constraints implied by Eq. (42) lead to $2n + 2$ conditions, that are linear equations/inequations.

Like related problems, the Knapsack problem, and the graph bipartitioning problem, we conjecture that POOR is also an *NP*-hard problem. An exact (brute force) solving strategy for POOR is to extract all point pairs satisfying the constraints, and then to select the pair which maximizes Eq. (46). This fixed point extraction referred to in Section 4.2 as the recognition process has been analyzed by Palm and Sommer (1992) to estimate the asymptotic information capacity bound of the optimal retrieval procedure. Of course, as a retrieval method this exact solution of POOR is computationally exhaustive, since it requires retrieval to be performed on all possible $\binom{n}{a}$ initial patterns with the wanted activity a. Fast algorithms, like the retrieval processes proposed in this paper, can only yield approximative solutions.

4.7. Approximative solutions of POOR

A possible approximative solving strategy of POOR consists of two phases. In a first phase the normalization constraints are released in order to find a pattern pair (X,Y) which contains the solution, i.e. $x^* \subset X$ and $y^* \subset Y$. In a second phase only subsets of (X,Y) are considered and the solution (x^*,y^*) is extracted by tightening the activity constraints.

The SB and the CB variant II retrieval strategies described in this paper follow this type of strategy. After the second retrieval step they have finished the first phase. The second phase has exclusively to eliminate wrong active components, and corresponds to the onset of the Boolean AND operation. With SB retrieval, the activity values in the pattern pair (X,Y) will be very high at high memory load and the second phase will often get stuck before reaching the desired low activity. The reason for the high activity in (X,Y) is that standard retrieval uses only half of the constraints implied by the fixed point condition Eq. (42) for the discrimination of active neurons — either row or column constraints. For high memory load, this discrimination can be too rough: even with the highest possible threshold — guaranteeing $|x(r)| \geq a$ — no activity reduction can be achieved. CB retrieval checks both row and column constraints in each retrieval step by forming the dendritic sum using the conditional links (see Section 3.3). This allows an activity reduction in cases where SB retrieval already fails to work.

5. Experiments

5.1. Retrieval errors and capacity

Both versions of CB retrieval have been tested in simulation experiments with random patterns, and compared with the standard retrieval model. Here, we show results for variant II with a parameter setting that has not been particularly optimized to maximum capacity. For more detailed experimental results with variant I, see Sommer and Palm (1998b) and Sommer et al. (1998). SB retrieval is not investigated experimentally, because it can be ruled out as a promising modification by the arguments in Section 4.4 for the parameter range we used.

Fig. 5(a) displays the experimental error rates in the retrieved y-patterns. For this parameter set, the synaptic matrix is filled with $p_1 = 0.38$. One-step retrieval produces unacceptably high mean errors as soon as initial noise is present, while CB retrieval achieves low error rates in the y-pattern and is much more robust to initial noise, even with

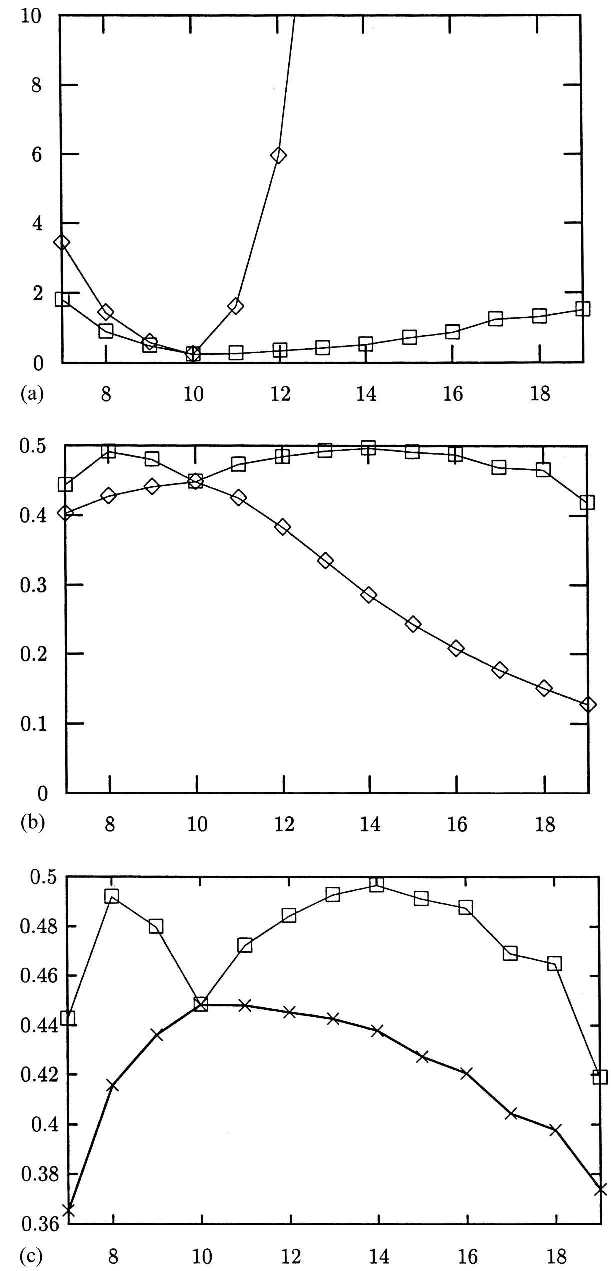

Fig. 5. Results of CB variant II (□) and the Willshaw model (◇) with $M = 20\,000$, but the other parameter setting and the x-axes labeled as in Fig. 2. Diagram (a) compares the expected number of 'false alarm' retrieval errors in the y-patterns. Diagram (b) displays the information efficiency. Due to retrieval errors, the capacity of the Willshaw model drops down very rapidly with increasing initial noise. For CB retrieval, diagram (c) displays output capacity and search capacity (= efficiency) in bit/synapse by (×) and (□), respectively. The difference between both curves is the contribution due to x-pattern completion, the completion capacity C. It is zero for $|x(0)| = 10$, if the initial pattern is error free.

such a high memory load. CB retrieval permits completion of the x-patterns with comparable quality. The total number of iteration steps depends on the initial pattern and with the proposed threshold strategy; it lies in a range between five and 50. The CB capacity values in Fig. 5(b) are close to the theoretical expectations given in Section 4.2. The capacity increases with slight distortions of the initial pattern due to the completion capacity contribution. It stays at a high level even if in the initial pattern every second active component is a 'false alarm' error. The one-step retrieval capacity is rapidly decreasing with initial errors. Because of the exploding retrieval errors of one-step retrieval [see Fig. 5(a)], its theoretical capacity cannot be fully exploited; in applications, the number of stored patterns has to be drastically reduced. Fig. 5(c) displays for the CB model how completion and output capacity contribute to the search capacity. For noiseless initial patterns, the pattern completion is zero and it assumes local maxima at certain values of 'miss' and 'add' errors.

5.2. Processing of ambiguous initial patterns

Retrieval from ambiguous initial patterns, where one memory pattern is superimposed not by random noise but by parts from one or a few other memory patterns, is a very likely scenario in nature as well as in many applications. Thus, a NAM model should be able to cope with such a situation.

Parts of several memory patterns can form an ambiguous initial pattern by two different kinds of superposition; either by Boolean AND or by the OR operation. The first kind produces an initial pattern with decreased but nonzero activity only in the special case of correlated patterns — with nonzero overlap. The more general kind of superposition is the second type, which produces an increased activity in the initial pattern. This case turned out as particularly hard for the standard model.

With CB retrieval, it is possible to process initial patterns containing an OR-superposition of several parts of stored patterns (see Sommer and Palm, 1998b). The predominating part will be singled out first. Its active components have to be deleted in the initial pattern to retrieve the next part and so on. Thus, a segmentation of the different initial components is achieved, and successively, a list of retrieved items ordered with respect to the relevance for the given initial pattern can be obtained. Only, if parts of several memory patterns with the same size are present in the initial pattern, the symmetry has to be broken before segmentation is possible in the CB model. This can be achieved by a random deletion process which singles out one predominate component in the initial pattern.

6. Conclusions

6.1. Implications of the analysis

In this paper, we have analyzed bidirectional retrieval from hetero-associative memories. The search capacity introduced in Section 4.1 describes the information balance of BAM, taking into account both pattern mapping and pattern completion. As a general performance measure for BAM models, we propose the information efficiency which is based on the search capacity, but also takes into account the required synaptic depth. For memories with binary synapses both quantities coincide. We calculate the asymptotic efficiency of the BAM Willshaw model in different cases of initial noise: $S(0,0) \leq 0.69$ bit/synapse and $S(0,0.5) \leq 0.52$ bit/synapse. For $S(0,0.5)$, we derive a theoretical bound of $\ln[2] = 0.69$, which is the universal efficiency bound of the BAM Willshaw model, valid for all parameter settings. This value agrees with the bound in the simple Willshaw model, indicating that BAM extension does not increase the asymptotic efficiency bound.

We present new methods for the theoretical description of the considered memory models for finite sizes. A finite size theory is important because the finite model behavior is entirely different from the asymptotic behavior and approaches the latter only for sizes far beyond realization. For instance, iterative retrieval schemes do improve finite models but are not required at all asymptotically (see Schwenker et al., 1996). The finite model cannot be described by simple elegant expressions as in the asymptotic case. The derived combinatorial formulae and recursive computation schemes expose their appeal not at first sight, but they allow easy and fast evaluation in a high level programming language like Mathematica. Our formula for the distribution of the dendritic sum of the Willshaw model (Proposition 4.1) is mathematically equivalent to a combinatorial expression given by Buckingham and Willshaw (1992), but in the interesting range of parameters the numerical evaluation is much simpler. The improved error analysis of the Willshaw model (Proposition 4.2) serves two purposes: (a) Proposition 4.3 derives the hit distribution of pattern part retrieval — a retrieval strategy in the original Willshaw model leading to an output pattern with limited activity, which is part of the proposed CB retrieval; and (b) Proposition 4.4 estimates the parameter range where standard bidirectional retrieval outperforms the Willshaw model. By theoretical analysis, standard bidirectional retrieval can be ruled out as a promising retrieval improvement (Section 4.4). Moreover, we derive a combinatorial error analysis of the first step of CB retrieval (Proposition 4.5) that corroborates its experimental superiority in the range where the performance of the finite Willshaw model is flawed by high retrieval errors, i.e. when memory load is high and activity in the initial pattern exceeds the activity of the memory patterns. The mathematical essence of Proposition 4.5 is the derivation of a recursive method to calculate the distribution of a sum of random integers.

We formulate a necessary condition for all memory patterns to be retrieved from the memory matrix and consider retrieval as an optimization problem. The problem with retrieving the appropriate memory pattern, given the memory matrix and the initial pattern, corresponds to a linear programming problem (POOR). We conjecture that POOR is *NP*-hard and explain how the considered retrieval strategies can be regarded as approximative solutions of POOR and what their differences are from this perspective.

6.2. The new BAM model

This paper suggests refined bidirectional retrieval methods in the hetero-associative Willshaw model, i.e. better exploitation of the information stored by the simple and incremental binary Hebbian learning process. In the Willshaw model, suppression of cross-talk noise has been proposed by specifically adjusted individual thresholds during one-step retrieval (Buckingham and Willshaw, 1993; Graham and Willshaw, 1995). However, this retrieval modification allows no completion of noisy addresses, and the computationally expensive threshold alignment cannot be accelerated by parallel hardware and requires additional information about the learning data. Our new retrieval strategy — CB retrieval — reduces cross-talk errors significantly without employing more complex learning procedures or dummy augmentation in the pattern coding as has been proposed to improve BAM (Wang et al., 1990; Leung et al., 1995). The conditional links, the quantities on which CB retrieval is essentially based, have a probabilistic interpretation: a high value between two units expresses a high probability that both units belong to the same memory pattern. This information can be exploited in various iteration schemes, two of which have been proposed and tested experimentally. CB retrieval yields an information efficiency of about 0.5 bit/synapse, which is close to the asymptotic value with low error rates even for initial patterns with high false alarm noise. It provides hetero-associative mapping, and can be used to complete noisy addresses and to segment superimposed addresses.

While the proposed NAM uses a very simple incremental learning strategy and quite sophisticated optimization retrieval strategies (type A), many of the reasonably efficient NAM models in the literature use sophisticated optimization learning strategies combined with simple retrieval (type B). In a type B memory, it takes longer to learn one new association the higher the number of stored associations already is, because in the learning process all previously learned patterns have to be presented several times. This is time consuming and it requires all learned patterns to be available, either by a kind of retrieval process during training, or from some kind of additional pattern store. A type A memory is more flexible in terms of learning, but if the number of stored associations grows, it will take longer to retrieve a learned association. As previously proposed auto-associative models have demonstrated (Gibson and Robinson, 1992; Hirase and Recce, 1996; Schwenker et al., 1996), and now the CB memory model shows for hetero-associative tasks, there are type A models that still retrieve efficiently and fast, since the required iteration numbers are low and operations required in an update step can be computed in parallel. In situations where retrieval time is critical, the iteration can also be interrupted at a preliminary result that can already be used as a reasonable approximation. The biological realization of associative memory may lie in between the two extremes discussed. We detail in Section 6.4 that refined iterative retrieval is suggested by the recurrent cortical connectivity. But also more elaborate learning has been proposed, for example as a functional model of REM sleep (Crick and Michison, 1986): in a one-step learning associative memory, a second learning phase provides an unsupervised reorganization of the information gathered by one-step learning. This REM sleep phase was shown in simulations to enhance the information capacity significantly, and to permit simpler retrieval (van Hemmen et al., 1990).

6.3. Application in information retrieval

Accessing records in large data bases according to user requests is the central problem of information retrieval. The application of the proposed NAM in information retrieval requires a similarity preserving coding of the data into sparse binary patterns. However, for user interaction, sparse representations meet natural preferences. Feature encoding, i.e. the extraction of feature sets which directly and quickly characterize complex situations, has been classified in cognitive psychology as one of the three basic types of cognitive processes (Sternberg, 1977). Such feature representations tend to be sparse: a description of 10 features out of 2000 possible features will be easier for a person to handle, compared with a description with only 90 possible features, but where around 50% of them apply. Since

$$\binom{90}{45} \simeq \binom{2000}{10},$$

the information contained in both descriptions is practically the same. The problem of finding appropriate sparse codings is application-dependent. For text indexing, word fragments used in existing indexing techniques (Gebhardt, 1987) can be directly used as sparse features. For image processing, the most natural features like lines and edges are usually sparse (Zetsche, 1990). Also a neural sparse coding model using anti-Hebbian learning has been proposed (Földiak, 1990). Sparse patterns extracted from different data channels in heterogeneous data can be easily combined (by concatenation) and processed simultaneously in the neural memory.

In information systems, the CB model offers an alternative to inverted indices, for the task of mapping from user queries to record locations in a similarity based, fault-tolerant manner. Improving the early suggestion of sparse associative memory for information retrieval by Bentz et al. (1989), our model offers the following advantages: (a) a user query should not only provide a data record, but also the completed feature description leading to the record (relevance feedback); and (b) ambiguous queries should not only trigger a single response, but a list of relevant records, ordered by their relevance (relevance ranking; see Section 5.2). A problem for image processing with sparse NAM models is the low information content of sparse patterns (Zetsche, 1990). It has been proposed to store larger items as sequences of sparse patterns (Kohring, 1990) which can be efficiently realized in the CB model: once the first pattern pair has been extracted by CB retrieval, unidirectional iteration of standard retrieval yields the pattern sequence.

6.4. Bidirectional retrieval in the cortex

The idea of BAM and the corresponding iterative retrieval strategies may be important not only in technical applications but also as a model of associative memory in the cortex. We end this paper with a few speculations in this direction. In the cortex, neighboring pyramidal cells have a high synaptic connection probability that, however, rapidly decreases with distance (Braitenberg and Schüz, 1991). Therefore, in small regions (cortical columns), iterative retrieval, such as in the Hopfield model, is likely, where each neuron performs several processing steps until the final retrieval pattern is obtained from the initial pattern (Amit, 1995). Many pyramidal cells also have long axon colaterals (Ramon y Cajal, 1911) that, as tracer studies have revealed, are organized in cortico–cortical pathways between pairs of different cortical areas. The reported pathways realize perhaps 20% of all possible direct connections between areas, and the majority of them are reciprocal, i.e. provide activity propagation in both directions (Felleman and Essen, 1991). Since many cortical areas can be assigned to particular tasks, the pathways must play a role in combining different kinds of information. A biological interpretation of CB retrieval (Sommer et al., 1998) examines the hypothesis that the 'expensive' reciprocal long range pathways provide efficient associative memory function. A detailed model with spiking neurons will be the subject of a forthcoming paper. It has already been shown that NAM with spiking neurons can efficiently process patterns of simultaneously spiking neurons (Wennekers and Sommer, 1998).

As a final remark, we briefly address the implications of BAM function in reciprocal projections of the cortex for the theory of Hebbian cell-assemblies as distributed multi-modal representations of concepts (Hebb, 1949). We assume that BAM operation of single pathways can be selectively enabled by input — and threshold — control. In such a cortex model, distributed cell-assemblies can emerge, if learned associations on different projections support each other. Cell-assemblies are formed by mutually supporting hetero-associative associations between pairs of cortical subrepresentations. In psychological terms, the basic memory units correspond to simple stimulus-reaction schemes, like habituation and conditioning, and multi-modal concept formation emerges as a secondary phenomenon. In this model, the information stored in a multi-modal cell-assembly can be accessed more flexibly without always activating the whole assembly: different parts of the assembly can be understood as different facets of a concept. This fits into the introspective observation that, depending on the situation, the recall of a learned concept may appear in various forms.

Acknowledgements

The authors would like to thank N. Palomero-Gallagher and T. Wennekers for their comments and suggestions which helped in improving the manuscript. This work was supported by grants SO352/3-1 and PA268/8-3 from the Deutsche Forschungsgemeinschaft.

References

Amit, D. J. (1995). The Hebbian paradigm reintegrated: local reverberations as internal representations. *Behav. Brain Sci.*, *18*, 617–657.

Bentz, H. J., Hagström, M., & Palm, G. (1989). Information storage and effective data retrieval in sparse matrices. *Neural Networks*, *2*, 289–293.

Braitenberg, V., & Schüz, A. (1991). *The anatomy of the cortex. Statistics and geometry*. New York: Springer.

Buckingham, J., & Willshaw, D. (1992). Performance characteristics of associative nets. *Network*, *3*, 407–414.

Buckingham, J., & Willshaw, D. (1993). On setting unit thresholds in an incompletely connected associative net. *Network*, *4*, 441–459.

Crick, F., & Michison, G. (1986). REM sleep and neural nets. *J. Mind Behav.*, *7*, 229–250.

Felleman, D. J., & Essen, D. C. V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

Földiak, P. (1990). Forming sparse representations by local anti-hebbian learning. *Biol. Cybern.*, *64*, 165–170.

Gardner-Medwin, A. (1976). The recall of events through the learning of associations between their parts. *Proc. R. Soc. Lond. B*, *194*, 375–402.

Gebhardt, F. (1987). Text signatures by superimposed coding of letter triplets and quadruplets. *Info. Syst.*, *12* (2), 151–156.

Gibson, W., & Robinson, J. (1992). Statistical analysis of the dynamics of a sparse associative memory. *Neural Networks*, *5*, 645–662.

Graham, B., & Willshaw, D. (1995). Improving recall from an associative memory. *Biol. Cybern.*, *72*, 337–346.

Haines, K., & Hecht-Nielsen, R. (1988). A BAM with increased information storage capacity. In *Proceedings of the International Conference on Neural Networks I* (pp. 181–190). Los Alamitos, CA: IEEE Press.

Hassoun, M. H. (1989). Dynamic heteroassociative neural memories. *Neural Networks*, *2*, 275–287.

Hebb, D. O. (1949). *The organization of behaviour*. New York: Wiley.

Hirase, H., & Recce, M. (1996). A search for the optimal thresholding sequence in an associative memory. *Network*, *7* (4), 741–756.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. In *Proceedings of the National Academy of Sciences, USA*, *79*, 2554–2558.

Kohonen, T. (1977). *Associative memory*. New York: Springer.

Kohring, G. A. (1990). Performance enhancement of Willshaw type networks through the use of limit cycles. *J. Phys. Fr.*, *51*, 2387–2393.

Kosko, B. (1987). Adaptive bidirectional associative memories. *Appl. Optics*, *26* (23), 4947–4971.

Leung, C.-S., Chan, L.-W., & Lai, E. (1995). Stability, capacity and statistical dynamics of second-order bidirectional associative memory. *IEEE Trans. Syst. Man Cybern.*, *25* (10), 1414–1424.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in neural activity. *Bul. Math. Biophys.*, *5*, 115–133.

Palm, G. (1980). On associative memory. *Biol. Cybern.*, *36*, 19–31.

Palm, G. (1988). On the asymptotic information storage capacity of neural networks. In R. Eckmiller, & C. v. d. Malsburg (Eds.), *Neural computers*. New York: Springer.

Palm, G. (1990). Cell assemblies as a guideline for brain research. *Concepts Neurosci.*, *1*, 133–148.

Palm, G., & Sommer, F. T. (1992). Information capacity in recurrent McCulloch–Pitts networks with sparsely coded memory states. *Network*, *3*, 1–10.

Palm, G., & Sommer, F. T. (1995). Associative data storage and retrieval in neural networks. In E. Domany, & J. L. van Hemmen, (Eds.), *Models of neural networks III* (pp. 79–118). New York: Springer.

Palm, G., Schwenker, F., Sommer, F. T., & Strey, A. (1997). Neural associative memory. In A. Krikelis, & C. Weems (Eds.), *Associative processing and processors.* Los Alamitos, CA: IEEE Press.

Potter, J. L. (1992). *Associative computing.* New York: Plenum Press.

Ramon y Cajal, S. (1911). *Histologie du systeme nerveux de l'homme et de vertebres*. Paris: Maloin.

Riordan, J. (1968). *Combinatorical identities*. New York: Wiley.

Schwenker, F., Sommer, F. T., & Palm, G. (1996). Iterative retrieval of sparsely coded associative memory patterns. *Neural Networks, 9* (3), 445–455.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Sommer, F. T., & Dayan, P. (1998). Bayesian retrieval in associative memories with storage errors. *IEEE Trans. Neural Networks*, *9* (4), 705–713.

Sommer, F. T., & Palm, G. (1998). Bidirectional retrieval from associative memory. In *Advances in Neural Information Processing Systems 10*. Cambridge, MA: MIT Press, pp. 675–681.

Sommer, F. T., Wennekers, T., & Palm, G. (1998). Bidirectional completion of Cell Assemblies in the cortex. In *Computational Neuroscience: Trends in Research*. New York: Plenum Press.

Steinbuch, K. (1961). Die Lernmatrix. *Kybernetik*, *1*, 36–45.

Sternberg, R. J. (1977). *Intelligence, information processing and analogical reasoning.* Hillsdale, NJ.

van Hemmen, J., Ioffe, L., Kühn, R., & Vaas, M. (1990). Increasing the efficiency of a neural network through unlearning. *Physika A, 163*, 386–392.

Wang, Y. F., Cruz, J. B., & Mulligan, J. H. (1990). Two coding stragegies for bidirectional associative memory. *IEEE Trans. Neural Networks, 1* (1), 81–92.

Wennekers, T., & Sommer, F. T. (1998). Gamma-oscillations support optimal retrieval in associative memories of two-compartment neurons. *Neurocomputing*, to appear.

Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Nonholographic associative memory. *Nature*, *222*, 960–962.

Zetsche, C. (1990). Sparse coding: the link between low level vision and associative memory. In R. Eckmiller, G. Hartmann, & G. Hauske (Eds.), *Parallel processing in neural systems and computers*. Amsterdam: North Holland/Elsevier.